Applicant Name: Brando Miranda
E-mail: miranda9@illinois.edu

Sketching: A Cognitively inspired Compositional Theorem Prover that Learns to Prove

Current advances in Deep Learning have accelerated A.I. in an unprecedented way. Now we can solve tasks that seemed beyond our reach like language translation, beating top human Go players and generating images indistinguishable from real ones. The progress is impressive but it has been at the cost of transparency. As models become larger and more complex it has become increasingly difficult to understand what they do reliably at many levels of abstraction. One level that inspires me is the mathematical level. Do current Deep Learning models generalize because of implicit regularization? Is the role of over-parameterization the factor that leads to learnability via gradient descent in highly non-convex landscapes? Are classical statistical learning bounds useful in characterizing Deep Neural Networks? With the increased investment in A.I. and people fearlessly building, designing and hacking, the race to explain what all these models do becomes unmanageably complex with our current mathematical tools.

In this proposal, we suggest that a promising way to tackle this problem is by building an A.I. system that learns and thinks intelligently as a human does about symbols and mathematics. This encompasses important problems like program induction, proof synthesis and theorem proving. There would be an immense amount of benefit for humanity by solving this problem. For example, every field that uses mathematical language would benefit. We would be able to solve questions about physics, mechanical engineering, computer science, machine learning, statistics, and many more. It's difficult to think of an application that wouldn't benefit from this but the most important one for us would be A.I. transparency and safety. A transparent and provable A.I. system is a safe and empowering system for humanity.

We propose that a promising way to tackle this problem is by leveraging the insights provided by human learning and cognitive science. One of the most powerful principles that can influence progress in mathematical reasoning as proof synthesis is the principle of Compositionality. The principle highlights that human knowledge, learning and reasoning are organized as reusable high-level concepts. We believe Compositionality is crucial in mathematical intelligence because it enables efficient reasoning by not worrying about every single detail immediately. If an intelligent system (artificial or not) computationally processes concepts blindly at a too low-level of granularity it may suffer from the famous combinatorial explosion problem because there are simply too many options. If however, it is able to identify the main and recurring high-level concepts, it can benefit from the abstraction of unimportant low-level details and thus reach a conclusion more efficiently. There is already evidence from theoretical Machine Learning that compositionality indeed might be one of the reasons Artificial Neural Networks learn more efficiently by overcoming the problem known as "the curse of dimensionality" [1].

Another observation from cognitive science is that humans are able to learn rich representations from fewer examples than current machine learning methods. The idea suggested is that this accelerated learning happens due to the principle of "learning-to-learn" [6] which states that previous experience aids in learning of future ideas. This is especially present in program induction via proof synthesis because Compositionality leverages the reusability of concepts. This leads to re-occurring concepts aiding in the incorporation of new concepts and

improved storage of these because of shared representations. This also avoids repetitive re-learning in the system and allows for greater power of generalization since it is able to use shared ideas in different scenarios. For example, mathematics is full of manipulation of logical symbols, quantifiers and implications. If the system is able to learn a flexible representation of these it inevitably benefits, because this is a widespread concept in all mathematics. Similarly, if the system has the capability of drawing relations between different concepts then when it detects this connection in new concepts it can more efficiently store it by explicitly representing this relation. For example, by noticing that the new concept is a special case of an old one it can re-use the previous and only incorporate any additional new properties (if present). Thus, its leverages the shared representations to more efficiently detect and store new concepts. A simple example in this domain would be abstract groups since many structures belong to this class of algebraic structures.

Despite Compositionality being a very promising idea, how can we concretely learn to use it for our A.I. systems to enjoy its benefits? How can a system automatically learn the contours and abstractions of mathematical concepts? Again we draw inspiration from how humans learn such concepts and notice that despite being widespread in how we educate people, the use of this idea is so far limited or non-existent in how we teach A.I. systems. One of the most powerful ways that humans learn is by using structured data (like textbooks, Wikipedia, etc.) that already outlines what the concepts are, and humans learn incrementally without having to discover everything from scratch. We suggest that new high-level concepts can be learned incrementally in the "curriculum learning paradigm" [2]. In the domain of proof synthesis, not only are there thousands of textbooks outlining the order, but there are also mathematical libraries with proofs already in program form ready to be used in this learning paradigm [3, 4, 5]. We suggest that this outline of the concepts are a source of semi-supervised data for the system to learn Compositionality of mathematics.

The last promising idea from cognitive science we will outline is Learning as Model Building. One of the reasons humans are so effective at learning programming and mathematics is because they learn beyond pattern recognition. They are able to build intuitive models that capture relationships and similarity between concepts. In other words, when people think about solving a new unseen problem they do not go through all the formal rules and "deduce the solution". Instead, we reason by analogies and predictions from the model we have built about the (real or mathematical) world. Those predictions simulate the world and outline the solution. Then if the predictions indeed arrive at the solution, we reinforce our current model or correct it. In the end, what we have is understanding, i.e. a model, capturing relationships and how one concept affects another.

In connection to the previous paragraph, we want to highlight an observation that this research direction might have surprising benefits that other fields of A.I. don't have, especially fields like Natural Language Processing (NLP). The advantage is that there exists an unambiguous reward signal of correctness. Contrast this to the difficult and important problem of defining sentence equivalence and semantics in natural language. This is difficult and there is no trivial solution to this. We believe this can be a powerful advantage as our system to guide its learning because truth is defined as provability in the domain.

In conclusion, we believe this research direction is crucial because any progress in it would benefit any human endeavour that uses mathematical thought. Furthermore, it yields a system that gives provable guarantees of other systems, which inevitably contributes to more transparent and safer A.I. Lastly, the approach we suggest drives A.I. development in a novel and promising direction that is currently underexplored.

1. T. Poggio, Mhaskar H., Rosasco L., Miranda B., and Liao Q., Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review, International Journal of Automation and Computing, pp. 1-17, 2017.
2. Yoshua Bengio, Jerome Louradour, Ronan Collobert, Jason Weston, Curriculum Learning.
3. Grzegorz Bancerek, Czesław Bylinski, Adam Grabowski, Artur Korniłowicz, Roman Matuszewski, Adam Naumowicz, Karol Pak, The Role of the Mizar Mathematical Library for Interactive Proof Development in Mizar.
4. Archive of Formal Proofs (in Isabelle), https://www.isa-afp.org/
5. Formalizing 100 Theorems, http://www.cs.ru.nl/~freek/100/index.html
6. Brenden M. Lake, Tomer D. Ullman, Joshua B. Tenenbaum, Samuel J. Gershman, Building Machines That Learn and Think Like People